社会的ジレンマ状況における効用関数による利得評価*

石　田　　　淳**

1．問題の所在

海野道郎（2006）は、社会現象を「社会的ジレンマ問題」として経験的に研究する際の基本指針として次の２点を指摘している。第１に、社会的ジレンマを人々の認知レベルではなく、実体レベルにおいて定義すべきであるということ。具体的にいえば、ゲームの「利得」を状況に対する行為者の評価の総体として捉えるのではなく、囚人のジレンマにおける刑期や金銭の大きさのように実体として捉える、ということである[1]。第２に、実際にそのような状況が存在し、結果として欠陥均衡が成立したという事実（「実現態」）によってではなく、潜在的な社会的ジレンマ構造が存在し、欠陥均衡が成立しえたという可能性（「可能態」）として、社会的ジレンマを定義すべきであるということ。つまり、社会的ジレンマは「実体レベル」かつ「可能態」として定義すべきであると海野は主張している。

この海野による枠組みに同意するならば、「実験状況や現実社会において、客観的には――実験状況の設定や社会現象の観察を行う研究者から見れば――社会的ジレンマ状況であるにもかかわらず、なぜ協力者が存在するのか？」というこれまでの経験的な社会的ジレンマ研究における主要な問いを簡潔な形で表すことができるだろう。それは、「社会的ジレンマ状況が理論的には『実現態』となるはずが、実際は『可能態』に留まるメカニズムはどのようなものか」というものである。そして、海野による「実体レベル」と「認知レベル」の区別を用いるならば、この問いに対するシンプルな解答を用意することができる。それは、「実体レベル」で見れば社会的ジレンマ状況であるにもかかわらず、人々の「認知レベル」では必ずしもそう認知されていない。ゆえに、各個人が主観合理的に行為選択をした場合、本来ならば現出するはずの「欠陥均衡」は現出しない、というものである。海野らの研究もこの方向性に沿って、人々の社会的ジレンマ状況に対する認知を分析している（海野 2006）。

本稿では、社会的ジレンマ状況に対する人々の認知の効果によってジレンマが「可能態」に留まるという一つの可能性を、限界効用逓減型効用関数による利得評価というメカニズムを想定することによって理論的に検討する。そのとき特に、本稿が注目するのは社会現象としての社会的ジレンマ問題がしばしばもつ「参与者の大規模性」という特性である。海野は、社会的ジレンマ状況の特性として、後に示す Dawes の２条件に対応するコスト感、危機感に加えて、参与者が多人数であることによって個人の影響力が小さくなるという感覚、つまり「無効性感覚」を挙げている。この無効性感覚は非協力行動を促進し欠陥均衡の実現を促すと考えられるが、本稿では逆に、ある特殊な条件のもとでは、集団が大規模であればあるほど協力の可能性が生まれるという結論が得られることになる。しかし、この結論を現実に適用する場合にはいくつかの留保が必要になることを最後に議論する。

---

*キーワード：社会的ジレンマ、効用関数、大規模集団
**日本学術振興会　特別研究員 PD

1）海野（1993）の指摘にもあるように、一般的にゲーム理論において利得は、フォン・ノイマン＝モルゲンシュテルン効用を意味し、期待利得＝期待効用の最大化を理論の中心に据えている。本稿では以下、海野による利得の「経験的解釈」に沿って、利得という概念によって「実体としての利得」を示すことにする。

2．効用関数による利得評価モデル

2.1.　社会的ジレンマ状況と状況の評価

以下、社会的ジレンマ状況と行為者による状況（利得構造）の評価をフォーマルに定式化しておこう。N 人のプレーヤーは、それぞれ C（協力）と D（非協力）という２つの選択肢をもつ。ここで、$D(m)$を m 人が C を選択しているときに D を選択したプレーヤーの受け取る（実体としての）利得、$C(m)$を自分自身を含む m 人が C を選択しているときに C を選択したプレーヤーが受け取る利得とする。ここで、$D(m)$、$C(m)$が次の Dawes の２つの条件（Dawes 1975，1980）に従うとき、それを「社会的ジレンマ状況」と呼ぶことにする。

すなわち、

〈条件１〉　$D(m) > C(m+1)$
　　$(0 \leq m \leq N-1)$

〈条件２〉　$C(n) > D(0)$

条件１を言い換えれば、任意の m について D を選択する誘因

$$t(m) = D(m) - C(m+1)$$

が常に正である、ということである。

ここで各プレーヤーがこの状況を認知した際、何らかの基準でその状況を評価すると仮定しよう。状況の評価は自らの利得だけではなく、相手の利得をも考慮する「利他的」な基準を考えることもできるだろう。ただし、ここでは自らの利得のみを考慮するという「利己的」な行為者を仮定する。その評価関数は一般的に $f: R \to R$ と定義される。個人 i の利得関数 $C(m)$の評価規則を f_i、$D(m)$の評価規則を g_i とする。そうすると、個人 i にとっての「認知上の社会的ジレンマ状況」は、$f_i(C(m))$、$g_i(D(m))$ であり、f_i、g_i の取り方によっては、

$$f_i(D(m)) > g_i(C(m+1))$$

が成立しない。ゆえに、実体レベルにおいては全員が非協力選択となるはずが、実際には認知上の評価の如何によっては、何人かの協力者が存在する可能性があるのである。このような説明の仕方は原理的に何通りものやり方を考えることができるだろう[2)]。以下では、アド・ホックではない評価規則の仮定の一つとして、限界効用逓減法則を満たす効用関数による利得の評価を取り上げる。

2.2.　効用関数の導入

効用関数による利得評価モデルとして、以下の仮定群を導入する。

仮定１．すべての個人は、$D(m)$、$C(m)$を評価する効用関数 U をもち、効用がより大きくなる選択肢を選択する。

仮定２．効用関数 U は、２階微分可能な連続関数であり、限界効用逓減法則を満たす。すなわち、$U' > 0$、$U'' < 0$。

仮定３．U の一階導関数は０に収束する。すなわち、$\lim_{x\to\infty} U'(x) = 0$。

仮定４．社会的ジレンマ状況における利得関数は Dawes の条件１、２に加えて次の条件を満たす。

〈条件３〉　$D'(m) > 0$、$\lim_{m\to\infty} D(m) = \infty$

〈条件４〉　$D(m) - C(m+1) = \alpha > 0$
　　$(0 \leq m \leq N-1)$

仮定１は先の認知的変換規則でいえば、$\forall i$、$f_i = g_i = U$ を意味している。仮定３は、一次の項を含むような限界効用逓減の度合いが低い関数を排除することを意味する。また、仮定４の条件３、４を満たす典型的な状況としては、$C(m) = am + b$、$D(m) = am + d$ という線形関数による定義がある。これは、木村邦博による分類の「非競合的集合財、コストは貢献者数にかかわらず一定、利益は貢献者数に比例」（木村 2002：73）タイプであり、典型的な N 人囚人のジレンマゲームである（図１）。

さて、仮定１から４を満たすとき、次の命題１、２が成立する。

命題１．利得上の社会的ジレンマが維持される限

2）経験的事象を説明するために実際には起こっておらず観察不可能な「可能態」を公理論的に設定する場合、厳密には「可能態」の設定の妥当性根拠が問われる必要がある。この問題は、FK モデル（高坂 2006）における「客観的階層」の問題と同種のものである（石田 2005）。

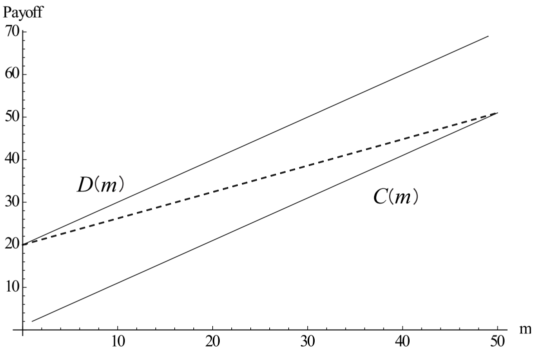

図１　社会的ジレンマ状況の例
（a＝1，b=1，d=20，N=50；点線は利得の平均）

り、効用関数 U 上においても Dawes の条件１、２は満たされる。
証明．仮定２の $U'>0$ より、

$$D(m)>C(m+1)\Rightarrow U(D(m))>U(C(m+1))$$
$$C(N)>D(0)\Rightarrow U(C(N))>U(D(0)) \qquad \square$$

U 上で Dawes の条件１、２が満たされるということは、認知レベルにおいても社会的ジレンマになっていることを示している。つまり、プレーヤーが主観合理的に選択したとしても全員非協力が実現するのである。

さてここで、$N\to\infty$を考えよう。社会的ジレンマ状況の参与者のオーダーが非常に大きくなる、という事態である。この場合、実体レベルで見た社会的ジレンマ状況はそのまま維持される。というのも $\lim_{m\to\infty}t(m)=\alpha>0$ だから条件１は満たされる。条件２が満たされることは自明である。しかし、効用関数による利得評価では事情は異なる。効用関数による利得評価上の誘因を $t_U(m)$ とする。

命題２．$N\to\infty$のとき、利得上の社会的ジレンマが維持される限り、$m=N$ となる m において、C から D へと選択を変更する効用関数上の誘因はゼロになる。つまり、$\lim_{m\to\infty}t_U(m)=0$ 。
証明．仮定４より、命題は

$$\lim_{m\to\infty}[U(D(m))-U(D(m)-\alpha)]=0$$

と言い換えることができる。ここで、U 上の点 $(D(m)、U(D(m)))$ と $(D(m)-\alpha、U(D(m)-\alpha))$ を通る直線を考える。この直線の傾きは

$$\frac{U(D(m))-U(D(m)-\alpha)}{\alpha}$$

である。ここで、平均値の定理より、任意の m について $D(m)-\alpha<c<D(m)$ かつ

$$U(D(m))-U(D(m)-\alpha)=\alpha U'(c) \qquad (1)$$

となる c が存在する。仮定４の〈条件３〉より、$m\to\infty\Rightarrow D(m)\to\infty$。また、$D(m)-\alpha<c<D(m)$ であるので $D(m)\to\infty\Rightarrow c\to\infty$ となる。ゆえに、

$$m\to\infty\Rightarrow c\to\infty \qquad (2)$$

が成立する。さらに仮定３より

$$\lim_{c\to\infty}\alpha U'(c)=0 \qquad (3)$$

である。結局、(1)、(2)、(3) より

$$\lim_{m\to\infty}[U(D(m))-U(D(m)-\alpha)]=\lim_{c\to\infty}\alpha U'(c)=0$$

が言える。 $\square$

命題２の特殊例として、効用関数としてしばしば用いられる $\ln x$ を考えてみると、$\lim(\ln x)'=\lim(1/x)=0$ となり、実際に t_U が０に近づいていくことを確認できる（図２、図３）。誘因がないときには、わざわざ非協力行動を取らないと解釈すると、$m=N$ のとき全員が C に留まることになる。また、ランダムにどちらかを選択すると考えた場合は、５割のプレーヤーが C に留まることになる。つまり、プレーヤー数が限りなく大きくなるとき、効用によってジレンマ状況を評価するプレーヤーを想定すると、協力の可能性が生じることがわかる。

命題２が本稿の主要命題であるが、今後の研究の拡がりのために次の命題を導入しておく。

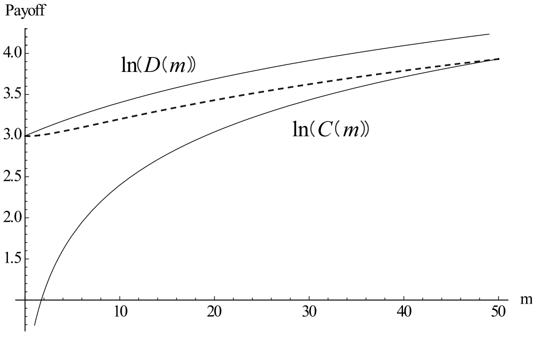

図２　社会的ジレンマ状況の ln による効用評価の例
（a＝1，b=1，d=20，N=50；点線は利得の平均）

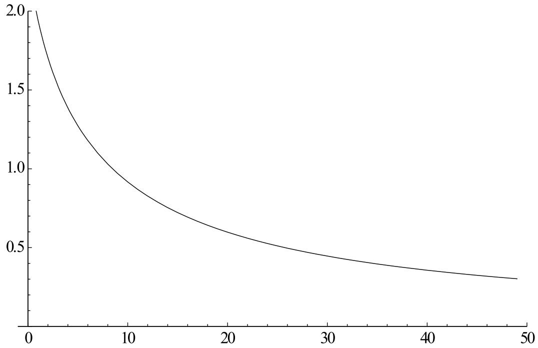

図３　$t_U(m)=\ln(D(m))-\ln(C(m+1))$ のグラフの例 ($a=1$, $b=1$, $d=20$, $N=50$)

命題３．仮定１、２、４が満たされるとき、$\lim_{x\to\infty}U'(x)=\beta\,(\beta\geq 0)$ となる効用関数 U について、$\lim_{m\to\infty}t_U(m)=\alpha\beta$ が成立する。
証明．命題２の証明において $\lim_{x\to\infty}U'(x)=\beta$ を仮定すればよい。□

３．結果の解釈とモデルの現実的な意味

ここまで、限界効用逓減型効用関数による利得評価というメカニズムを導入することで、大規模社会的ジレンマ状況において協力者が存在する可能性を検討してきた。その結果、プレーヤー数が無限に限りなく近づくとき、全員協力を所与としたとき協力者が積極的に非協力へと転じる誘因が限りなく小さくなることが分かった。このことは、社会的ジレンマが「可能態」に留まる１つのメカニズムと見なすことができるかもしれない。しかしながら、本稿のモデルの現実的な意味を考える際には、いくつかの注釈が必要になる。最後に３点に触れておこう。

3.1.　Olson 問題との関連

N が大きくなると協力者の存在の可能性が出てくる、という結論は、一見すると M. Olson (1965＝1983) の提起した Olson 問題と矛盾している[3)]。しかし、この見た目の矛盾はモデルの帰結を以下のように解釈することによって説明することができる。

モデルの理論的な帰結であるナッシュ均衡解は同一であっても、実際の現象に照らして考えると、社会的ジレンマ状況において $m=0$ を出発点として各プレーヤーが協力に移行するかどうかを選択する場合——典型的には公共財供給問題——と、$m=N$ を出発点として各プレーヤーが非協力に移行するかどうかを選択する場合——典型的には資源管理問題——とがある。それぞれのケースについて、本稿のモデルの帰結を考えよう。まず $m=0$ を出発点とする場合である。命題１より、$m=0$ では全員が D に留まる。ゆえに、全員非協力となり公共財供給は始動されない。これは N の大きさに依存しない結論である。一方、$m=N$ からスタートする場合を考えると、N が有限の範囲においては命題１より全員が D へと移行する。しかし、$N\to\infty$ のとき、C と D は無差別になる。ゆえに、協力選択者が存在する可能性があり、その場合全員非協力には至らない。

Olson 問題は主に公共財供給に関するものであり、このように考えると、本モデルの結論と矛盾しないことが分かる。ただし、Olson の知見を積極的に支持するわけでもない。

ところで、非協力標準型ゲームをプレーヤー集合が無限集合であるケースまで拡張して定義した場合、先に示した $U(C(m))$、$U(D(m))$ で利得が得られるなら、効用変換モデルにおいて $N\to\infty$ のとき、$m=0$ と $m=N$ はナッシュ均衡になる。このとき、$m=0$ はパレート最適ではなく、$m=N$ はパレート最適である[4)]。

3.2.　利得のタイプと効用

本稿のモデルでは、利得のすべての部分を同一

3) ここで Olson 問題を、木村（2002）にならって、「集団規模が大きくなればなるほど、社会的に最適な集合財の供給状態が達成されにくくなること」と定義しておく。

4) Dawes の条件１と命題１より、$\forall i\in\{1,2,...,N\}, U(D(0))>U(C(1))$ なので、$m=0$ がナッシュ均衡。また、条件２と命題１より、$\forall i\in\{1,2,...,N\}, U(D(0))<U(C(N))$ となるので、$m=0$ はパレート最適ではない。命題２より、$N\to\infty$ かつ $m=N$ のとき、すべてのプレーヤーにとって C から D への単独の移行は効用を増大させない。また、U の単調増加性より、任意の他の状態への移行による効用の変化は、あるプレーヤーが C に移行するか、D に移行するかにかかわらず $U(x)-U(x-\beta)$（ただし β は定数で $0<\beta<\infty$）の正負を考えればよい。ところで、これも命題２と同様に考えると、$\lim_{x\to\infty}[U(x)-U(x-\beta)]=0$。つまり、有限の数のプレーヤーが D 選択へ移行した場合、すべてのプレーヤーの効用はまったく変化しないのである。このことから実は、

の効用関数によって評価する、と想定していた．しかしながら、より現実的な想定として、金銭や時間など限界効用逓減が生じにくい利得部分と、サービスや利便性といった限界効用逓減が生じやすい利得部分に分かれると考えることもできる。２つの場合について考えてみよう。

利得を線形関数 $C(m)=am+b$、$D(m)=am+d$ の形で定式化した社会的ジレンマ状況は、公共財供給問題を考えた場合、木村の定式化にもあるように、「公共財から得る利得（利益）」am と「コストを払わないことによる利得（コスト回避利得）」$d-b$ に分けることができる。ここで、公共財から得る利得のみが仮定２、３を満たす効用関数 u によって評価され、コスト回避利得はそのままの値が評価値に加算されると想定しよう。効用上の誘因は $t_U(m)=u(am)-u(am-a)+d-b$ である。この場合、命題２から、$\lim_{m\to\infty}t_U(m)=d-b$ となる。

また、利得が一定の比率で非金銭的利得と、コスト回避利得を含む金銭的利得に分かれており、非金銭的利得が効用評価において仮定２、３を満たす効用関数 u によって評価され、金銭的利得は線形関数によって評価されると想定することもできる。ここで、$D(m)=x$ として、非金銭的利得部分を πx、金銭的利得部分を $(1-\pi)x$ とおく（ただし $0\leq\pi\leq 1$）。このとき、効用による評価値は $U(x)=u(\pi x)+\beta(1-\pi)x$ である。この場合、$\lim_{x\to\infty}U'(x)=\beta(1-\pi)$ となるので、命題３を適用することで、$\lim_{m\to\infty}t_U(m)=\alpha\beta(1-\pi)$ と結論づけられる。

このように、部分的に限界効用逓減を想定した場合には、命題２は成立せず、一概に N が大きくなると協力者の存在の可能性が出てくるとは言えなくなる。ただし、上述のどちらのケースでも、効用関数による利得評価上の誘因 $t_U(m)$ は、m について単調減少する。

3.3. 無限大の参与者と無限大の利得

最後に当然のことながら、N を無限にすることの実質的な意味を考える必要がある。数学上の意味での無限は現実には存在しない。N が有限だとすると、どれだけ誘因が小さくなろうと、合理的選択は D である。逆に、それほどの「厳密な合理的選択」が果たして可能か、ということを問う必要がある。認知的ヒューリスティックスを考慮に入れると、モデルの帰結は「N が大きいとき、プレーヤーは D への誘因をほとんど感じなくなる」くらいの弱い傾向性を示していると考えるべきであろう。

さらに、利得がプレーヤー数について単調増加し無限大に発散する、という状況の現実性を考える必要がある。本稿の結論は、非常に膨大な参与者が、それぞれ非常に膨大な利得を享受しているときの話である。しかしながら、資源や財の有限性ということを考慮すれば、現実にそのようなケースがどれほどありうるだろうか？　このように考えると、本稿のモデルの現実的妥当性は結局のところ低いものであり、「実現することのないユートピアとしての全員協力」を示唆しているものと位置づけられる。つまり、旧来のジレンマ・モデルの結論の頑健性を別の観点から示す結果になっているのである。いずれにせよ、効用関数による利得評価というメカニズムのさらなる検討によって、社会的ジレンマ・メカニズムに対する理解がさらに深まることを期待している。

【付記】

本稿は、平成18～20年度科学研究費補助金（特別研究員奨励費）による研究成果の一部である。

参考文献

Dawes, Robyn M., 1975, “Formal Models of Dilemma in Social Decision-making,” pp. 87–107 in Martin F. Kaplan and Steven Schwartz (eds.), *Human Judgment and Decision Processes*, New York:

非協力者数が有限の整数であるような戦略の組み合わせは、すべてナッシュ均衡であることが導かれる。ナッシュ均衡を無限プレーヤー集合について定義すると、効用関数の取り方によってしばしばこのような非直感的な結論が導かれる。無限プレーヤー集合を仮定したゲームは、ゲーム理論では non-atomic game として分析されている。例えば、Schmeidler（1973）は他者の平均的反応に依存する特殊な効用関数を仮定したとき、non-atomic game において純粋戦略ナッシュ均衡が存在することを示している。

Academic Press.

————————, 1980, “Social Dilemmas,” *Annual Review of Psychology*, 31：169–93.

石田淳，2005,「経験科学としての数理社会学――FK モデルにおける客観階層システムとモデル検証問題」『理論と方法』20(1)：97–108.

木村邦博，2002,『大集団のジレンマ――集合行為と集団規模の数理』ミネルヴァ書房.

髙坂健次，2006,『社会学におけるフォーマル・セオリー――階層イメージに関する FK モデル』ハーベスト社.

Olson, Mancur, 1965, *The Logic of Collective Action: Public Goods and the Theory of Groups*, Cambridge: Harvard University Press.（＝依田博・森脇俊雅（訳），1983,『集合行為論――公共財と集団理論』ミネルヴァ書房.

Schmeidler, David, 1973, “Equilibrium Points of Nonatomic Games,” *Journal of Statistical Physics*, 7(4)：295–300.

海野道郎，1993,「合理的選択理論の基礎概念――社会的ジレンマ研究を支える諸概念に関する検討」海野道郎（編）『社会的ジレンマに関する数理社会学的研究』（平成３～４年度科学研究費補助金（総合研究 A）研究成果報告書）：1–18.

————，2006,「誰が社会的ジレンマ状況を定義するのか？――社会的ジレンマ状況の定義と人々の行動」『社会学研究』80：7–28.

A Model of Evaluation of Payoffs by Utility Functions in a Social Dilemma Situation

ABSTRACT

The aim of this paper is to make a theoretical study on the possibility of the existence of cooperators in large-scale social dilemma situation. For this purpose, a model of evaluation of payoffs by utility function is introduced. The analysis portion of this model shows that the larger the player set becomes, there is lesser incentive for cooperators to deceive one another when all of the player are cooperators. In the end, an infinitely large player set results in disappearance of incentive. This model can be regarded as one possible mechanism that evades the realization of social dilemma. However, in order to come to this conclusion, it is first necessary to assume an unreal situation.

Key Words: social dilemma, utility function, large-scale group